

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-175696

(43)Date of publication of application : 24.06.1994

(51)Int.Cl. G10L 9/14  
G10L 3/00  
G10L 3/00  
G10L 9/18

(21)Application number : 05-201795

(71)Applicant : INTERNATL BUSINESS MACH  
CORP <IBM>

(22)Date of filing : 22.07.1993

(72)Inventor : BAHL LALIT R  
DE SOUZA PETER V  
GOPALAKRISHNAN PONANI S  
PICHENY MICHAEL A

(30)Priority

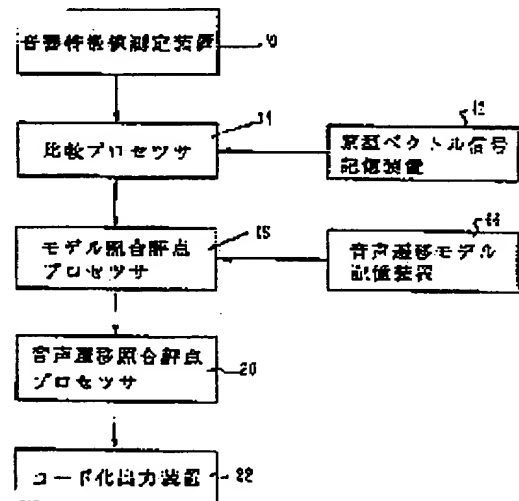
Priority number : 92 942862 Priority date : 10.09.1992 Priority country : US

(54) VOICE CODER, METHOD THEREFOR, VOICE RECOGNIZING DEVICE AND METHOD THEREFOR

(57)Abstract:

PURPOSE: To provide a voice coder and a method for collating fast acoustic.

CONSTITUTION: A voice coder compares how close the feature value of a feature vector of one pronunciation to the parameter value of a model vector signal to obtain plural model collating marks concerning the feature vector and each model vector and to store plural voice transition models expressing voice transition. A model collating mark concerning a first feature vector and each voice transition model includes output probability concerning at least one model collating marks concerning a first feature vector signal and a model collating mark. A voice transition collating mark concerning the first feature vector signal and each voice transition includes a best model collating mark concerning all the voice transition models expressing voice transition and the first feature vector. The identification value of each voice transition and the voice transition collating mark concerning the first feature vector signal and the each voice transition are outputted as a pronunciation expressing signal obtained by coding the first feature vector signal.



## LEGAL STATUS

[Date of request for examination]

22.07.1993

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2986313

[Date of registration] 01.10.1999

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19)日本国特許庁(J P)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-175696

(43)公開日 平成6年(1994)6月24日

(51)Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 9/14		G 8946-5H		
		J 8946-5H		
3/00	5 3 1	D 9379-5H		
	5 3 5	9379-5H		
9/18		E 8946-5H		

審査請求 有 請求項の数9(全 21 頁)

(21)出願番号 特願平5-201795

(22)出願日 平成5年(1993)7月22日

(31)優先権主張番号 07/942862

(32)優先日 1992年9月10日

(33)優先権主張国 米国(US)

(71)出願人 390009531

インターナショナル・ビジネス・マシー  
ズ・コーポレーション

INTERNATIONAL BUSIN  
ESS MACHINES CORPO  
RATION

アメリカ合衆国10504、ニューヨーク州  
アーモンク (番地なし)

(74)代理人 弁理士 合田 潔 (外2名)

最終頁に続く

(54)【発明の名称】 音声コード化装置及びその方法並びに音声認識装置及びその方法

(57)【要約】

【目的】高速音響照合のための音声コード化装置及びその方法を提案する。

【構成】音声コード化装置は原型ベクトル信号のパラメータ値に対する1つの発音の特徴ベクトル信号の特徴値の近さを比較することにより特徴ベクトル信号及び各原型ベクトル信号についての複数の原型照合評点を得、音声遷移を表す複数の音声遷移モデルを記憶する。第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含む。第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点は音声遷移を表すすべての音声遷移モデルと第1の特徴ベクトル信号とについての最良のモデル照合評点を含む。各音声遷移の識別値と第1特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とは第1の特徴ベクトル信号のコード化された発音表現信号として出力される。

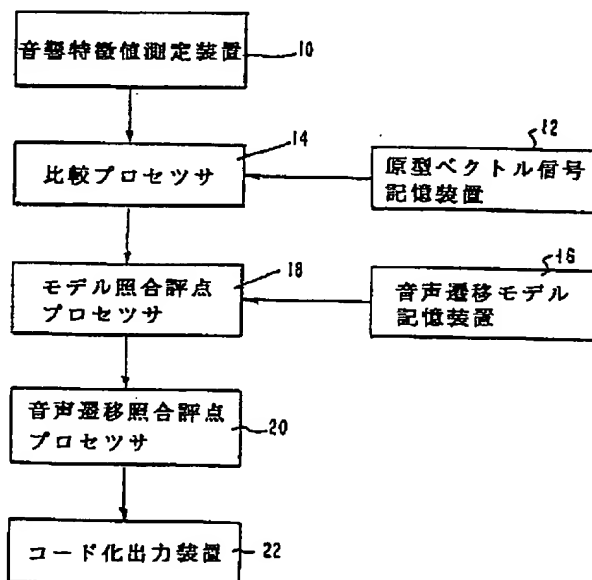


図1 本発明の音声コード化装置

## 【特許請求の範囲】

【請求項1】一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより上記特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶する手段と、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する手段と、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する手段とを具え、少なくとも1つの音声遷移は複数の異なるモデルによって表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有し、各モデル照合評点は第1の特徴ベクトル及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、各音声遷移照合評点は第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むことを特徴とする音声コード化装置。

【請求項2】さらに、識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶する手段と、第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する手段とを具え、各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含み、各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含み、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する上記手段は、各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力することを特徴とする請求項1に記載の音声コード化装置。

【請求項3】一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより上記特徴値を表す一連の特徴ベクトル信号を発生するステップと、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶するステップと、原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得るステップと、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶し、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力に関する出力確率を有するようにするステップと、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成し、各モデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含むようにするステップと、第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成し、各音声遷移照合評点は第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むようにするステップと、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力するステップとを含むことを特徴とする音声コード化方法。

【請求項4】さらに、識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶し、各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含むようにするステップと、第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成し、各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含むようにするステップとを含み、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する上記出力ステップは、各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力することを特徴とする請求項3に記載の音声コード化方法。

【請求項5】一連の連続的な各時間間隔について発音の

少なくとも1つの特徴の値を測定することにより上記特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶する手段と、各特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、各特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する手段と、識別値を有する音声ユニットをそれぞれ表す複数の音声ユニットモデルを記憶する手段と、各特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する手段と、各音声ユニットの識別値と特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを特徴ベクトル信号のコード化した発音表現信号として出力する手段と、それぞれが少なくとも1つの音声ユニットモデルを含む複数の語についての確率論的モデルを記憶する手段と、一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成する手段と、最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別する手段と、少なくとも1つの最良の候補語を出力する手段とを具え、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力に関する出力確率を有し、特徴ベクトル信号についてのモデル照合評点は特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、特徴ベクトル信号についての音声遷移照合評点は特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含み、音声ユニットを表す各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含み、特徴ベクトル信号についての音声ユニット照合評点は特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含み、各語モデルは開始状態と、終了状態と、上記開始状態から上記終了状態への道の少なくとも一部で音声ユニットモデルを通る複数の経路とを有し、

各語照合評点は一連の特徴ベクトル信号と語のモデル内の一連の音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニットとについての音声ユニット照合評点の組合せを含むことを特徴とする音声認識装置。

【請求項6】原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る上記手段は、各特徴ベクトル信号に対する各原型ベクトル信号の推定した近さの順序に原型ベクトル信号をランク付けすることにより各特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を得、特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点は特徴ベクトル信号及び原型ベクトル信号についてのランク評点を含むことを特徴とする請求項5に記載の音声認識装置。

【請求項7】一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより上記特徴値を表す一連の特徴ベクトル信号を発生するステップと、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶するステップと、原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得るステップと、識別値を有する音声遷移でなる語彙からの音声遷移を表す複数の音声遷移モデルを記憶し、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有するようにするステップと、各特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成し、特徴ベクトル信号についてのモデル照合評点は特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含むようにするステップと、各特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成し、特徴ベクトル信号についての音声遷移照合評点は特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むようにするステップと、識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶し、各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含むようにするステップと、各特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成し、特徴ベクトル信号についての音声ユニット照合評点は特徴ベクトル信号と音声ユニ

ット内のすべての音声遷移とについての最良の音声遷移照合評点を含むようにするステップと、

各音声ユニットの識別値と特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを特徴ベクトル信号のコード化した発音表現信号として出力するステップと、

それぞれが少なくとも1つの音声ユニットモデルを含む複数の語についての確率論的モデルを記憶し、各語モデルは開始状態と、終了状態と、上記開始状態から上記終了状態への道の少なくとも一部で音声ユニットモデルを通る複数の経路とを有するようにするステップと、

一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成し、各語照合評点は一連の特徴ベクトル信号と語のモデル内の一連の音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニットとについての音声ユニット照合評点の組合せを含むようにするステップと、

最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別するステップと、

少なくとも1つの最良の候補語を出力するステップとを含むことを特徴とする音声認識方法。

【請求項8】原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る上記ステップは、各特徴ベクトル信号に対する各原型ベクトル信号の推定した近さの順序に原型ベクトル信号をランク付けすることにより各特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を得、

特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点は特徴ベクトル信号及び原型ベクトル信号についてのランク評点を含むことを特徴とする請求項7に記載の音声認識方法。

【請求項9】一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより上記特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、

原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、

識別値を有する音声遷移でなる語彙からの音声遷移を表す複数の音声遷移モデルを記憶する手段と、

第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、

識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶する手段と、

第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する手段とを具え、

少なくとも1つの音声遷移は複数の異なるモデルによって表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有し、

各モデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、

各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含み、

各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内の音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含み、

出力手段は各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力することを特徴とする音声コード化装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は音声コード化装置及びその方法並びに音声認識装置及びその方法に関し、例えば音声認識システム等のための音声コード化装置及び方法に適用して好適なものである。

【0002】

【従来の技術】音声認識システムにおいて文脈非依存又は文脈依存の音響モデルを使用して語、音素、音素の一部の発音をモデル化することは知られている。文脈依存音響モデルは前後に発音される語又は語の一部に頼つて語又は語の一部の発音をシミュレートする。従つて文脈依存音響モデルは文脈非依存音響モデルよりも一段と正確である。しかしながら文脈依存音響モデルを使用する発音認識は、文脈非依存音響モデルを用いる発音認識に比べ一段と多くの演算を必要とし、従つて一段と多くの時間を必要とする。

【0003】また音声認識システムにおいて、高速音響照合により候補語の選択候補名簿を迅速に選択した後、高速音響照合で選択した各候補語を詳細音響照合によつて一段と注意深く評価することも知られている。候補語を迅速に選択するために高速音響照合に文脈非依存音響モデルを使用することも知られている。高速音響照合によつて選択した各候補語を一段と注意深く評価するために詳細音響照合に文脈依存音響モデルを使用することも知られている。

【0004】

【発明が解決しようとする課題】本発明の目的は詳細音響照合に用いられるものと同じ文脈依存音響モデルを使用する、高速音響照合用の音声コード化装置及び音声コード化方法を提供することである。

【0005】本発明の他の目的は詳細音響照合に用いら

れるものと同じ文脈依存音響モデルを使用する、高速音響照合を有する音声認識装置及び音声認識方法を提供することである。

【0006】

【課題を解決するための手段】かかる課題を解決するため本発明においては、一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶する手段と、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する手段と、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する手段とを設け、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有し、各モデル照合評点は第1の特徴ベクトル及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、各音声遷移照合評点は第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むようにする。

【0007】また本発明においては、一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、特徴値を表す一連の特徴ベクトル信号を発生するステップと、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶するステップと、原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得るステップと、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶し、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力に関する出力確率を有するようにするステップと、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成し、各モデル照合評点は第1の特徴ベクトル信号及び原型ベク

トル信号についての少なくとも1つの原型照合評点に関する出力確率を含むようにするステップと、第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成し、各音声遷移照合評点は第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むようにするステップと、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力するステップとを含むようにする。

【0008】また本発明においては、一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、識別値を有する音声遷移でなる語彙からの音声遷移をそれぞれ表す複数の音声遷移モデルを記憶する手段と、各特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、各特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する手段と、識別値を有する音声ユニットをそれぞれ表す複数の音声ユニットモデルを記憶する手段と、各特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する手段と、各音声ユニットの識別値と特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを特徴ベクトル信号のコード化した発音表現信号として出力する手段と、それぞれが少なくとも1つの音声ユニットモデルを含む複数の語についての確率論的モデルを記憶する手段と、一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成する手段と、最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別する手段と、少なくとも1つの最良の候補語を出力する手段とを設け、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力に関する出力確率を有し、特徴ベクトル信号についてのモデル照合評点は特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、特徴ベクトル信号についての音声遷移照合評点は特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含み、音声ユニットを表す各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含み、特徴ベクトル信号についての音声ユニット照合評点は特徴ベクトル信号と音声ユニット内のすべての音声遷

移とについての最良の音声遷移照合評点を含み、各語モデルは開始状態と、終了状態と、開始状態から終了状態への道の少なくとも一部で音声ユニットモデルを通る複数の経路とを有し、各語照合評点は一連の特徴ベクトル信号と語のモデル内の一連の音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニットとについての音声ユニット照合評点の組合せを含むようにする。

【0009】また本発明においては、一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、特徴値を表す一連の特徴ベクトル信号を発生するステップと、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶するステップと、原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得るステップと、識別値を有する音声遷移でなる語彙からの音声遷移を表す複数の音声遷移モデルを記憶し、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有するようにするステップと、各特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成し、特徴ベクトル信号についてのモデル照合評点は特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含むようにするステップと、各特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成し、特徴ベクトル信号についての音声遷移照合評点は特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含むようにするステップと、識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶し、各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含むようにするステップと、各特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成し、特徴ベクトル信号についての音声ユニット照合評点は特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含むようにするステップと、各音声ユニットの識別値と特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを特徴ベクトル信号のコード化した発音表現信号として出力するステップと、それぞれが少なくとも1つの音声ユニットモデルを含む複数の語についての確率論的モデルを記憶し、各語モデルは開始状態と、終了状態と、開始状態から終了状態への道の少なくとも一部で音声ユニットモデルを通る複数の経路とを有するようにするステップと、一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成し、各語照合評点は一連の特徴ベクトル信号と語のモデル内の一連の

音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニットとについての音声ユニット照合評点の組合せを含むようにするステップと、最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別するステップと、少なくとも1つの最良の候補語を出力するステップとを含むようにする。

【0010】さらに本発明においては、一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、特徴値を表す一連の特徴ベクトル信号を発生する手段と、それぞれが少なくとも1つのパラメータ値を有する複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る手段と、識別値を有する音声遷移でなる語彙からの音声遷移を表す複数の音声遷移モデルを記憶する手段と、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成する手段と、識別値を有する音声ユニットを表す複数の音声ユニットモデルを記憶する手段と、第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する手段とを設け、少なくとも1つの音声遷移は複数の異なるモデルによつて表され、各音声遷移モデルは複数のモデル出力を有し、各モデル出力は原型ベクトル信号についての原型照合評点を含み、各音声遷移モデルは各モデル出力についての出力確率を有し、各モデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含み、各音声ユニットモデルは2つ又は3つ以上の音声遷移及び2つ又は3つ以上の音声遷移モデルを含み、各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内の音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含み、出力手段は各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力するようにする。

【0011】

【作用】本発明の音声コード化装置は以下のような機能を有する手段を含む。一連の各連続的な時間間隔について発音の少なくとも1つの特徴の値を測定することにより、この特徴値を表す一連の特徴ベクトル信号を発生する手段を含む。複数の原型ベクトル信号を記憶する記憶手段を含む。各原型ベクトル信号は少なくとも1つのパラメータ値を有する。原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより、第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る比較手段を含む。

【0012】複数の音声遷移モデルを記憶する記憶手段



を含む。各音声遷移モデルは音声遷移でなる語彙からの音声遷移を表す。各音声遷移は識別値を有する。複数の異なるモデルによつて少なくとも1つの音声遷移が表現される。各音声遷移モデルは複数のモデル出力を有する。各モデル出力は原型ベクトル信号についての原型照合評点を含む。また各音声遷移モデルは各モデル出力についての出力確率も有する。

【0013】第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成するモデル照合評点手段を含む。各モデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含む。

【0014】第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する音声遷移照合評点手段を含む。各音声遷移照合評点は第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含む。

【0015】最後に、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する出力手段を含む。

【0016】さらに本発明による音声コード化装置は以下のような機能を有する手段を含む。複数の音声ユニットモデルを記憶する記憶手段を含む。各音声ユニットモデルは2つ又は3つ以上の音声遷移を含む音声ユニットを表す。各音声ユニットモデルは2つ又は3つ以上の音声遷移モデルを含む。各音声ユニットは識別値を有する。

【0017】第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する音声ユニット照合評点手段を含む。各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含む。

【0018】本発明のこうした特徴において、上述の出力手段は各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する。

【0019】上述の比較手段は、例えば第1の特徴ベクトル信号に対する各原型ベクトル信号の推定した近さの順序に原型ベクトル信号をランク付けすることにより、第1の特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を得る。この場合、第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点は第1の特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を含む。

【0020】各音声遷移モデルは前後の音声遷移の独特の文脈内に対応する音声遷移を表すのが好適である。各音声ユニットは音素であるのが好適であり、各音声遷移は音素の一部であるのが好適である。

【0021】本発明による音声認識装置は、一連の各連続的な時間間隔について発音の少なくとも1つの特徴の値を測定することによりその特徴値を表す一連の特徴ベクトル信号を発生する手段と、複数の原型ベクトル信号を記憶する手段と、原型ベクトル信号のパラメータ値に対する各特徴ベクトル信号の特徴値の近さを比較することにより各特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る比較手段と、複数の音声遷移モデルを記憶する記憶手段と、各特徴ベクトル信号と各音声遷移モデルとについてのモデル照合評点を生成するモデル照合評点手段と、各特徴ベクトル信号及び各音声遷移についての音声遷移照合評点をモデル照合評点から生成する音声遷移照合評点手段と、2つ又は3つ以上の音声遷移モデルを含む複数の音声ユニットモデルを記憶する手段と、各特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を音声遷移照合評点から生成する音声ユニット照合評点手段と、各音声ユニットの識別値と1つの特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを特徴ベクトル信号のコード化した発音表現信号として出力する手段とを含む。

【0022】さらに本発明の音声認識装置は以下のような機能を有する手段を含む。複数の語についての確率論的モデルを記憶する記憶手段を含む。各語モデルは少なくとも1つの音声ユニットモデルを含む。各語モデルは開始状態と、終了状態と、開始状態から終了状態への道の少なくとも一部で音声ユニットモデルを通る複数の経路とを有する。一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成する語照合評点手段を含む。各語照合評点は一連の特徴ベクトル信号と語のモデル内の一連の音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニットとについての音声ユニット照合評点の組合せを含む。最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別する最良候補手段を含む。少なくとも1つの最良の候補語を出力する出力手段を含む。

【0023】本発明に従つて各音声遷移についての照合評点としてその音声遷移のすべてのモデルについての最良の照合評点を選択することにより、音声コード化及び音声認識のための装置及び方法は詳細音響照合において使用しているものと同じ文脈依存音響モデルを高速音響照合に使用することができる。

【0024】

【実施例】以下図面について、本発明の一実施例を詳述する。

【0025】図1は本発明による音声コード化装置の一例を示すブロック図である。この音声コード化装置は一連の連続的な各時間間隔について発音の少なくとも1つの特徴の値を測定することにより、その特徴値を表す一連の特徴ベクトル信号を発生する音響特徴値測定装置1

0を含む。次表(1)は間隔をおいた時間(t)1、2、3、4及び5にそれぞれ対応する仮想の一連の一次元特徴ベクトル信号を示す。

【0026】

【表1】

表1

時間 ( t )	特徴ベクトル F V ( t )
1	0 . 7 9 2
2	0 . 0 5 4
3	0 . 6 3
4	0 . 4 3 4
5	0 . 4 3 8

【0027】以下に詳細に説明するように、この時間間隔は10[msec]ごとに取り持続時間20[msec]のサンプルであるのが好ましい。

【0028】さらに音声コード化装置は複数の原型ベクトル信号を記憶する原型ベクトル信号記憶装置12を含む。各原型ベクトル信号は少なくとも1つのパラメータ値を有する。

【0029】次表(2)はそれぞれが1つのパラメータ値を有する、9つの仮想の原型ベクトル信号の例PV1a、PV1b、PV1c、PV2a、PV2b、PV3a、PV3b、PV3c及びPV3dを示す。

【0030】

【表2】

表2

原型ベクトル信号	パラメータ値	F V ( 1 ) への近さ	2進の原型照合評点	個別ラング ク原型照合評点	グループ ランク原型照合評点
P V 1 a	0 . 0 4 2	0 . 7 5 0	0	8	3
P V 1 b	0 . 4 8 3	0 . 3 0 9	0	3	3
P V 1 c	0 . 0 4 9	0 . 7 4 3	0	7	3
P V 2 a	0 . 7 6 9	0 . 0 2 3	1	1	1
P V 2 b	0 . 9 5 7	0 . 1 6 5	0	2	2
P V 3 a	0 . 4 3 3	0 . 3 5 9	0	4	3
P V 3 b	0 . 3 0 0	0 . 4 9 2	0	6	3
P V 3 c	0 . 4 0 8	0 . 3 8 4	0	5	3
P V 3 d	0 . 0 0 2	0 . 7 9 0	0	9	3

【0031】比較プロセッサ14は原型ベクトル信号のパラメータ値に対する第1の特徴ベクトル信号の特徴値の近さを比較することにより、第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点を得る。

【0032】上記の表(2)は原型ベクトル信号のパラメータ値に対する表(1)の特徴ベクトルF V (1)の近さの仮想の例を示す。この仮想の例に示すように、原型ベクトル信号P V 2 aは特徴ベクトル信号F V (1)に最も近い原型ベクトル信号である。最も近い原型ベク

トル信号に対する原型照合評点が「1」に定義され、他のすべての原型ベクトル信号に対する原型照合評点が「0」に定義されると、「2進」の原型照合評点「1」が原型ベクトル信号P V 2 aに対して割り当てられる。他のすべての原型ベクトル信号には「2進」の原型照合評点「0」が割り当てられる。

【0033】またこれ以外の原型照合評点を使用してもよい。例えば比較プロセッサ14は、第1の特徴ベクトル信号に対する各原型ベクトル信号の推定近さの順序に複数の原型ベクトル信号をランク付けすることにより第

1の特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を得るランク付け手段を含んでもよい。これにより、第1の特徴ベクトル信号及び各原型ベクトル信号についての原型照合評点は第1の特徴ベクトル信号及び各原型ベクトル信号についてのランク評点を含む。

【0034】表(2)は「2進」原型照合評点に加えて、個別ランク原型照合評点及び群ランク原型照合評点の例を示す。

【0035】この仮定の例においては、特徴ベクトル信号及び原型ベクトル信号が1つの次元だけを有し、この次元についてのただ1つのパラメータ値だけを有するものとして示した。しかしながら実際には、特徴ベクトル信号及び原型ベクトル信号が例えば50の次元を有することがある。各次元は各原型ベクトル信号について2つのパラメータ値を有することができる。各次元の2つのパラメータ値は、例えば平均値及び標準偏差(分散)値であつてよい。

【0036】再度図1を参照する。さらに音声コード化装置は複数の音声遷移モデルを記憶する音声遷移モデル記憶装置16を含む。各音声遷移モデルは音声遷移でなる語彙からの音声遷移を表す。各音声遷移は1つの識別値を有する。複数の異なるモデルによつて少なくとも1つの音声遷移が表される。各音声遷移モデルは複数のモデル出力を有する。各モデル出力は原型ベクトル信号についての原型照合評点を含む。各音声遷移モデルは各モデル出力についての出力確率を有する。

【0037】次表(3)は仮定の例として3つの音声遷移ST1、ST2及びST3を示す。これらは複数の異なる音声遷移モデルによつて表される。音声遷移ST1は音声遷移モデルTM1及びTM3によつてモデル化される。音声遷移ST2は音声遷移モデルTM4、TM5、TM6、TM7及びTM8によつてモデル化され

る。音声遷移ST3は音声遷移モデルTM9及びTM10によつてモデル化される。

【0038】

【表3】

表3

音声遷移 識別値	音声遷移 モデル
ST1	TM1
ST1	TM2
ST1	TM3
ST2	TM4
ST2	TM5
ST2	TM6
ST2	TM7
ST2	TM8
ST3	TM9
ST3	TM10

【0039】次表(4)は音声遷移モデルTM1～TM10の仮定の例を示す。この仮定の例の中の各音声遷移モデルはゼロでない出力確率を有する2つのモデル出力を含む。各出力は1つの原型ベクトル信号についての原型照合評点を含む。他のすべての原型ベクトル信号についてのすべての原型照合評点はゼロの出力確率を有する。

【0040】

【表4】

表 4

モデル出力				モデル出力			
音声遷移 モデル	原型ベク トル信号	原型照合 評点	出力 確率	原型ベク トル信号	原型照合 評点	出力 確率	
TM 1	P V 3 d	1	0.511	P V 3 c	1	0.489	
TM 2	P V 1 b	1	0.636	P V 1 a	1	0.364	
TM 3	P V 2 b	1	0.682	P V 2 a	1	0.318	
TM 4	P V 1 a	1	0.975	P V 1 b	1	0.025	
TM 5	P V 1 c	1	0.899	P V 1 b	1	0.101	
TM 6	P V 3 d	1	0.566	P V 3 c	1	0.434	
TM 7	P V 2 b	1	0.848	P V 2 a	1	0.152	
TM 8	P V 1 b	1	0.994	P V 1 a	1	0.006	
TM 9	P V 3 c	1	0.178	P V 3 b	1	0.822	
TM 10	P V 1 b	1	0.384	P V 1 a	1	0.616	

【0041】記憶しておく音声遷移モデルは、例えばマルコフモデル又は他の動的プログラミングモデルであつてよい。音声遷移モデルのパラメータは、例えば順方向一逆方向アルゴリズムにより得られる平滑化パラメータによつて、周知の発音訓練テキストから推定してよい（例えば、1976年4月発行、IEEE会報、第64巻、第4号「統計的手法による連続的音声認識」第532～536頁を参照）。

【0042】各音声遷移モデルは前後の音声遷移又は音素の独特の文脈の中の対応する音声遷移を表す。文脈依存音声遷移モデルは、例えば先ず文脈非依存モデルを構成することによつて作成することができる。これは、音素のモデルから手動によつて構成してもよく、例えば米国特許第4,759,068号「複数の音声から語のマルコフモデルを構成する方法」に述べられている方法によつて自動的に構成してもよく、また文脈非依存モデルを作成する他の周知の方法によつて構成してもよい。

【0043】次に音声遷移の発音を文脈依存語彙にグループ化することによつて文脈依存モデルを作成することができる。文脈を手動で選択することができるし、又は

音声遷移に対応する各特徴ベクトル信号にその文脈をもつてタグを付け、この特徴ベクトル信号をこれらの文脈に従つてグループ化することにより選択した評価関数を最適化することによつて自動的に選択することもできる。

【0044】再び図1を参照する。さらに音声コード化装置は、第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点を生成するモデル照合評点プロセッサ18を含む。各モデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含む。

【0045】次表(5)は表(2)の2進原型照合評点を使用して、特徴ベクトル信号FV(1)及び表(4)に示した各音声遷移モデルについてのモデル照合評点の仮定の例を示す。表(4)に示すように、2進原型照合評点が「1」である原型ベクトル信号PV2aの出力確率はTM3及びTM7以外のすべての音声遷移モデルについてはゼロである。

【0046】

【表5】

表5

音声遷移識別値	音声遷移モデル	FV(1) についてのモデル照合評点
ST1	TM1	0
ST1	TM2	0
ST1	TM3	0.318
ST2	TM4	0
ST2	TM5	0
ST2	TM6	0
ST2	TM7	0.152
ST2	TM8	0
ST3	TM9	0
ST3	TM10	0

【0047】さらに音声コード化装置は音声遷移照合評点プロセッサ20を含む。音声遷移照合評点プロセッサ20は第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点を生成する。各音声遷移照合評点は、第1の特徴ベクトル信号と音声遷移を表すすべての音声遷移モデルとについての最良のモデル照合評点を含む。

【0048】次表(6)は特徴ベクトル信号FV(1)及び各音声遷移についての音声遷移照合評点の仮想の例を示す。表(5)に示すように、特徴ベクトル信号FV

(1)及び音声遷移ST1についての最良のモデル照合評点は、音声遷移モデルTM3に対する0.318のモデル照合評点である。特徴ベクトル信号FV(1)及び音声遷移ST2についての最良のモデル照合評点は音声遷移モデルTM7に対する0.152のモデル照合評点である。同様に特徴ベクトル信号FV(1)及び音声遷移ST3についての最良のモデル照合評点はゼロである。

【0049】

【表6】

表6

音声遷移識別値	FV(1) についての音声遷移照合評点
ST1	0.318
ST2	0.152
ST3	0

【0050】図1に示す音声コード化装置は、各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力するコード化出力手段22を含む。表(6)は特徴ベクトル信号FV(1)に対するコード化出力の仮想の例を示す。

【0051】図2は本発明による音声コード化装置の他の例を示すブロック図である。この実施例における音響特徴値測定装置10、原型ベクトル信号記憶装置12、比較プロセッサ14、モデル照合評点プロセッサ18及び音声遷移照合評点プロセッサ20は図1に関連して述べた要素と同じである。しかしながら、この実施例において音声コード化装置は複数の音声ユニットモデルを記

憶する音声ユニットモデル記憶装置24を含む。各音声ユニットモデルは2つ又は3つ以上の音声遷移を含む音声ユニットを表す。各音声ユニットモデルは2つ又は3つ以上の音声遷移モデルを含む。各音声ユニットは識別値を有する。各音声ユニットは音素であり、各音声遷移は音素の一部であるのが好ましい。

【0052】次表(7)は音声ユニット(音素)P1及びP2にそれぞれ対応する音声ユニットモデルSU1及びSU2の仮想の例を示す。音声ユニットP1は音声遷移ST1及び音声遷移ST3を含む。音声ユニットP2は音声遷移ST2及び音声遷移ST3を含む。

【0053】

【表7】

表 7

音声ユニット 識別値	音声ユニット モデル	音声ユニットの中 の音声遷移		F V ( 1 ) についての 音声ユニット照合評点
P 1	S U 1	S T 1	S T 3	0 . 3 1 8
P 2	S U 2	S T 2	S T 3	0 . 1 5 2

【0054】再度図2を参照すると、さらに音声コード化装置は音声ユニット照合評点プロセッサ26を含む。音声ユニット照合評点プロセッサ26は第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点を生成する。各音声ユニット照合評点は第1の特徴ベクトル信号と音声ユニット内のすべての音声遷移とについての最良の音声遷移照合評点を含む。

【0055】本発明による音声コード化装置のこの実施例において、コード化出力手段22は各音声ユニットの識別値と第1の特徴ベクトル信号及び各音声ユニットについての音声ユニット照合評点とを第1の特徴ベクトル信号のコード化した発音表現信号として出力する。

【0056】表(7)の仮想の例に示すように、特徴ベクトル信号F V ( 1 ) についてのコード化した発音表現信号は、音声ユニットP 1及びP 2についての識別値とそれぞれ0.318及び0.152の音声ユニット照合評点とを含む。

【0057】図3は本発明の音声コード化装置を用いる本発明による音声認識装置の一実施例のブロック図である。この音声認識装置は音声コード化装置28を含み、音声コード化装置28は図2に示すすべての要素を含む。さらに音声認識装置は複数の語についての確率論的モデルを記憶する語モデル記憶装置30を含む。各語モデルは少なくとも1つの音声ユニットモデルを含む。各語モデルは開始状態と、終了状態と、開始状態から終了状態への道の少なくとも一部に音声ユニットモデルを通る複数の経路とを有する。

【0058】図4は語又は語の一部についての音響モデルの仮想の例を示す。図4に示す仮想モデルは開始状態

S 1、終了状態S 4及び開始状態S 1から終了状態S 4への道の少なくとも一部に複数の経路を含む。図4に示す仮想モデルは音声ユニットのモデルP 1、P 2及びP 3を含む。

【0059】図5は音素についての音響モデルの仮想の例を示す。この例において、音響モデルは遷移T 1の3つの発生セグメント、遷移T 2の4つの発生セグメント及び遷移T 3の3つの発生セグメントを含む。破線で示す発生セグメントは空白遷移である。実線で示した各遷移は原型ベクトル信号についての原型照合評点を含むモデル出力を有する音声遷移モデルでモデル化される。各モデル出力は1つの出力確率を有する。各空白遷移は出力をもっていない遷移モデルでモデル化される。

【0060】上述の手法で各語の複数の発音から自動的に語モデルを構成するか又は音声モデルから手動で構成しても良い。

【0061】図3に戻り、さらに音声認識装置は語照合評点プロセッサ32を含む。語照合評点プロセッサ32は一連の特徴ベクトル信号及び複数の各語についての語照合評点を生成する。各語照合評点は一連の音声ユニットモデルを通る少なくとも1つの経路に沿う音声ユニット及び一連の特徴ベクトル信号についての音声ユニット照合評点と語のモデルとの組合せを含む。

【0062】次表(8)は特徴ベクトルF V ( 1 )、F V ( 2 )及びF V ( 3 )と音声ユニットP 1、P 2及びP 3とについての音声ユニット照合評点の仮想の例を示す。

【0063】

【表8】

表 8

音声ユニット	F V ( 1 ) についての音声ユニット照合評点	F V ( 2 ) についての音声ユニット照合評点	F V ( 3 ) についての音声ユニット照合評点
P 1	0 . 3 1 8	0 . 2 0 4	0 . 8 2 5
P 2	0 . 1 5 2	0 . 9 7 9	0 . 7 0 7
P 3	0 . 4 3 9	0 . 6 3 5	0 . 2 7 3

【0064】次表(9)は図4に示す仮想の音響モデルの遷移についての遷移確率の仮想の例を示す。

【0065】

【表9】

表 9

音声ユニット	遷移	遷移確率
P 1	S 1 → S 1	0. 2
P 1	S 1 → S 2	0. 8
P 2	S 2 → S 2	0. 3
P 2	S 2 → S 3	0. 7
P 3	S 3 → S 3	0. 2
P 3	S 3 → S 4	0. 8

【0066】次表(10)は図4の音響モデルの各遷移についての特徴ベクトルFV(1)、FV(2)及びFV(3)の遷移確率の仮定の例を示す。

【0067】  
【表10】

表 10

開始状態	次の状態	FV(1)の確率	FV(2)の確率	FV(3)の確率
S 1	S 1	0. 0636	0. 0408	0. 165
S 1	S 2	0. 2544	0. 1632	0. 66
S 2	S 2	0. 0456	0. 2937	0. 2121
S 2	S 3	0. 1064	0. 6853	0. 4949
S 3	S 3	0. 0878	0. 127	0. 0546
S 3	S 4	0. 3512	0. 508	0. 2184

【0068】図6は図4の音響モデルを通る経路と表(8)、表(9)及び表(10)の仮定パラメータを使用するこのモデル及び一連の特徴ベクトル信号についての語照合評点の生成との仮定の例を示す。図6において変数Pは各ノードへの到達の確率(すなわち、各時間における各状態への到達の確率)である。

【0069】図3に戻り、さらに音声認識装置は最良の語照合評点を有する1つ又は2つ以上の最良の候補語を識別する最良候補語識別装置34を含む。語出力装置36は少なくとも1つの最良の候補語を出力する。

【0070】専用又は汎用のデジタルコンピュータシステムのいずれかを適切にプログラミングすることによって本発明による音声コード化装置及び音声認識装置を作ることができる。さらに詳細に述べれば、専用又は汎用のデジタルプロセッサのいずれかを適切にプログラミングすることによって比較プロセッサ14、モデル照合評点プロセッサ18、音声遷移照合評点プロセッサ20、音声ユニット照合評点プロセッサ26、語照合評点プロセッサ32及び最良候補語識別装置34を作ることができる。原型ベクトル信号記憶装置12、音声遷移モデル記憶装置16、音声ユニットモデル記憶装置24及び語モデル記憶装置30は電子計算機のメモリであつてよい。語出力装置36は、例えば陰極線管、液晶表示装置又はプリンタ等のビデオ表示装置であつてよい。また

語出力装置36はスピーカ又はヘッドホンを含む音声合成装置等のオーディオ出力装置であつてもよい。

【0071】図7は音響特徴値測定装置の一例を示す。この測定手段は発音に対応するアナログ電気信号を発生するマイクロホン38を含む。マイクロホン38のアナログ電気信号はアナログ-デジタル変換器40によってデジタル電気信号に変換される。この目的を達成するために、アナログ-デジタル変換器40が例えば20[kHz]のレートでアナログ信号をサンプリングする。

【0072】ウインドウ発生器42は、例えばアナログ-デジタル変換器40からの持続時間20[msec]のデジタル信号のサンプルを10[msec](1センチ秒)ごとに得る。持続時間20[msec]のデジタル信号の各サンプルをスペクトル分析器44によって分析することにより、例えば20の各周波数帯におけるこのデジタル信号のサンプルの大きさを得る。またスペクトル分析器44はこの20[msec]のデジタル信号サンプルの合計の大きさをすなわちトータルパワーを表わす21次元信号を発生するのが好適である。スペクトル分析器44は例えば高速フーリエ変換プロセッサであつてよい。またスペクトル分析器44は20個の帯域フィルタのバンクであつてもよい。

【0073】適応性を有する雑音除去プロセッサ46によって背景雑音を除去するようにスペクトル分析器44

が発生する21次元ベクトル信号を適合させることができる。雑音消去プロセッサ46は雑音消去プロセッサ内に与えられる特徴ベクトル $F(t)$ 入力から雑音ベクトル $N(t)$ を減算することにより、特徴ベクトル $F'(t)$

( $t$ )を出力として発生する。雑音消去プロセッサ46は前の特徴ベクトル $F(t-1)$ を雑音又は無音である

$$N(t) = \frac{N(t-1) + k[F(t-1) - F_p(t-1)]}{(1+k)}$$

【0075】ここで、 $N(t)$ は時間( $t$ )における雑音ベクトル、 $N(t-1)$ は時間( $t-1$ )における雑音ベクトル、 $k$ は適応性雑音消去モデルの固定パラメータ、 $F(t-1)$ は時間( $t-1$ )における雑音消去プロセッサ46内への特徴ベクトル入力であり雑音又は無音を表し、 $F_p(t-1)$ は特徴ベクトル $F(t-1)$ に最も近い、記憶装置48からの1つの無音又は雑音原型ベクトルである。

【0076】(A)ベクトルの総エネルギーがスレシヨルドを下回るか又は(B)適応性原型ベクトル記憶装置50内の、特徴ベクトルに最も近い原型ベクトルが雑音又は無音を表す原型のとき、前の特徴ベクトル $F(t-1)$ は雑音又は無音として認識される。特徴ベクトルの総エネルギーを分析するこうした目的のためには、スレシヨルドは、例えば特徴ベクトルが評価される前の2秒

$$X_i(t) = F'_i(t) - Z(t)$$

【0079】ここで $F'(t)$ は正規化されていないベクトルの時間( $t$ )における $i$ 番目の成分、 $Z(t)$ は $F'(t)$ の成分及び次式(3)及び次式(4)による

$$Z(t) = 0.9Z(t-1) + 0.1M(t)$$

【0081】

$$M(t) = \frac{1}{20} \sum_i F'_i(t)$$

【0082】正規化された20次元特徴ベクトル $X(t)$ を適応性ラベラ54でさらに処理することにより音声の発音の変動に適合させることができる。適応性ラベラ54の入力端に供給された20次元特徴ベクトル $X(t)$ から20次元適応ベクトル $A(t)$ を減算することによつ

$$A(t) = \frac{A(t-1) + k[X(t-1) - X_p(t-1)]}{(1+k)}$$

【0084】ここで $k$ は適応性ラベリングモデルの固定パラメータ、 $X(t-1)$ は時間( $t-1$ )における適応性ラベラ54への正規化された20次元ベクトル入力、 $X_p(t-1)$ は時間( $t-1$ )における20次元特徴ベクトル $X(t-1)$ に最も近い(適応性原型記憶装置50からの)適応性原型ベクトル、 $A(t-1)$ は時間( $t-1$ )における適応性ベクトルである。

【0085】適応性ラベラ54からの20次元の適合され

と識別したならばいつでも、雑音ベクトル $N(t)$ を定期的に更新することによつて雑音レベルを変更するように適合される。雑音ベクトル $N(t)$ を次式(1)に従つて更新する。

【0074】

【数1】

..... (1)

間で生成された(音声及び無音の双方に対応する)すべての特徴ベクトルの5パーセント点であつてよい。

【0077】雑音消去の後、特徴ベクトル $F'(t)$ を正規化することにより短期間平均正規化プロセッサ52によつて入力音声の大きさの変動を調整する。正規化プロセッサ52は21次元特徴ベクトル $F'(t)$ を正規化することにより20次元の正規化された特徴ベクトル $X(t)$

( $t$ )を発生する。合計の大きさをなわちトータルパワーを表す、特徴ベクトル $F'(t)$ の21次元目は放棄される。正規化された特徴ベクトル $X(t)$ の時間 $t$ における各成分 $i$ は、例えば次式(2)によつて対数ドメインで与えられる。

【0078】

【数2】

..... (2)

$Z(t-1)$ の加重平均である。

【0080】

【数3】

..... (3)

【数4】

..... (4)

て、適合された20次元特徴ベクトル $X'(t)$ が発生される。時間( $t$ )における適応性ベクトル $A(t)$ は例えば次式(5)から得ることができる。

【0083】

【数5】

..... (5)

た特徴ベクトル $X'(t)$ は聴覚モデル56に供給される。例えば、聴覚モデル56は人の聴覚システムがいかにして音響信号を知覚するかモデルを提供する。米国特許第4,980,918号「音韻論的グラフの効率的記憶及び高速アセンブリを伴う音声認識システム」に聴覚モデルの一例が述べられている。

【0086】本発明によると、聴覚モデル56は時間( $t$ )における適合された特徴ベクトル信号 $X'(t)$



の各周波数帯  $i$  について、次式 (6) 及び次式 (7) に従って新しいパラメータ  $E_i(t)$  を算出する。

$$E_i(t) = K_1 + K_2(X'_i(t))(N_i(t-1))$$

【0088】

$$N_i(t) = K_3 \times N_i(t-1) - E_i(t-1)$$

【0089】ここで、 $K_1$ 、 $K_2$  及び  $K_3$  は聴覚モデルの固定パラメータである。

【0090】センチ秒の各時間間隔について聴覚モデル56の出力は修正された20次元の特徴ベクトル信号である。この特徴ベクトルには、他の20の次元の値を2乗したものの合計の平方根に等しい値を有する21番目の次元が付加される。

【0091】センチ秒の各時間間隔について結合器58は、1つの現在のセンチ秒時間間隔と、4つの先行するセンチ秒時間間隔と、4つの後続のセンチ秒時間間隔とを表す9個の21次元特徴ベクトルを結合することにより、単一の189次元の接続されたベクトルを形成する。ロータ60で各189次元の接続した結合ベクトルに回転行列を掛けることにより、この接続したベクトルを回転させてこの接続したベクトルを50次元に減らす。

【0092】訓練セッションの期間中に得られる189次元の接続したベクトルのセットを例えばM個のクラスに分類することによって、ロータ60で使用する回転行列を得ることができる。訓練セット内のすべての接続したベクトルについての共分散行列に、M個のすべてのクラス内における接続したベクトルのすべてについてのクラス内の共分散行列の逆を掛ける。その結果得られる行列の最初の50個の固有ベクトルが回転行列を形成する。

(一例として、1989年12月発行、IBMテクニカル・ディスクロージャ・ブリテン、第32巻、第7号「離散パラメータ音素に基づいたマルコフ語モデルを使用する音声認識システム用のベクトル量子化手順」320頁及び321頁参照)。

【0093】ウインドウ発生器42、スペクトル分析器44、適応性を有する雑音除去プロセッサ46、短時間平均正規化プロセッサ52、適応性を有するラベラ54、聴覚モデル56、結合器58及びロータ60は適切にプログラムされた専用又は汎用のデジタル信号プロセッサであつてよい。原型記憶装置48及び50は上述の形式の電子計算機のメモリであつてよい。

【0094】例えば、訓練セットからの特徴ベクトル信号を複数のクラスにクラスタ化し、次に各クラスについての平均偏差及び標準偏差を算出して原型ベクトルのパラメータ値を形成することによって、原型記憶装置48の原型ベクトルを得ることができる。訓練用のスクリプトが一連の語セグメントモデル(一連の語のモデルを形成する)を含み、各語セグメントモデルが語セグメントモデル内に特定の記憶場所を有する一連の基本モデル

【0087】

【数6】

..... (6)

【数7】

..... (7)

ルを含むとき、各クラスが単一の語セグメントモデル内の1つの記憶場所の単一の基本モデルに対応することを指定することによって、特徴ベクトル信号をクラスタ化することができる。この方法は、1991年7月16日出願、米国特許出願第730,714号「自動音声認識用の音響原型を引き出す高速アルゴリズム」に一段と詳細に述べられている。

【0095】また訓練用テキストの発音によって生成され、所与の基本モデルに対応するすべての音響特徴ベクトルを、K-平均ユークリッドクラスタ化又はK-平均ガウスクラスタ化又はこれらの双方によってクラスタ化することもできる。この方法の一例は、1991年3月22日出願、米国特許出願第673,810号「話者非依存型のラベルコード化装置」に述べられている。

【0096】

【発明の効果】上述のように本発明によれば、音声コード化装置は、原型ベクトル信号のパラメータ値に対する発音の特徴ベクトル信号の特徴値の近さを比較することにより特徴ベクトル信号及び各原型ベクトル信号についての複数の原型照合評点を得、音声遷移を表す複数の音声遷移モデルを記憶する。第1の特徴ベクトル信号及び各音声遷移モデルについてのモデル照合評点は第1の特徴ベクトル信号及び原型ベクトル信号についての少なくとも1つの原型照合評点に関する出力確率を含む。第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点は音声遷移を表すすべての音声遷移モデルと第1の特徴ベクトル信号とについての最良のモデル照合評点を含む。各音声遷移の識別値と第1の特徴ベクトル信号及び各音声遷移についての音声遷移照合評点とは第1の特徴ベクトル信号のコード化した発音表現信号として出力される。これにより詳細音響照合で使用されるものと同じ文脈依存音響モデルを使用する、高速音響照合用の音声コード化装置を提供することができる。

【図面の簡単な説明】

【図1】図1は本発明による音声コード化装置の一実施例を示すブロック図である。

【図2】図2は本発明による音声コード化装置の他の実施例を示すブロック図である。

【図3】図3は本発明による音声コード化装置を用いた本発明の音声認識装置の一実施例を示すブロック図である。

【図4】図4は語又は語についての一部の音響モデルの仮想の例を示す略線図である。

【図5】図5は音素についての音響モデルの仮定の例を示す略線図である。

【図6】図6は図4の音響モデルを通る完全な経路及び部分的な経路の仮定の例を示す略線図である。

【図7】図7は本発明による音声コード化装置及び音声認識装置に使用される音響特徴値測定装置の一実施例を示すブロック図である。

【符号の説明】

10……音響特徴値測定装置、12……原型ベクトル信号記憶装置、14……比較プロセッサ、16……音声遷移モデル記憶装置、18……モデル照合評点プロセッサ、20……音声遷移照合評点プロセッサ、22……コード

化出力手段、24……音声ユニットモデル記憶装置、26……音声ユニット照合評点プロセッサ、28……音声コード化装置、30……語モデル記憶装置、32……語照合評点プロセッサ、34……最良候補語識別装置、36……語出力装置、38……マイクロホン、40……アナログ-デジタル変換器、42……ウインドウ発生器、44……スペクトル分析器、46……適応性を有する雑音消去プロセッサ、48……無音又は雑音原型ベクトル記憶装置、50……適応性原型ベクトル記憶装置、52……平均正規化プロセッサ、54……適応性を有するラベラ、56……聴覚モデル、58……結合器、60……ロテータ。

【図1】

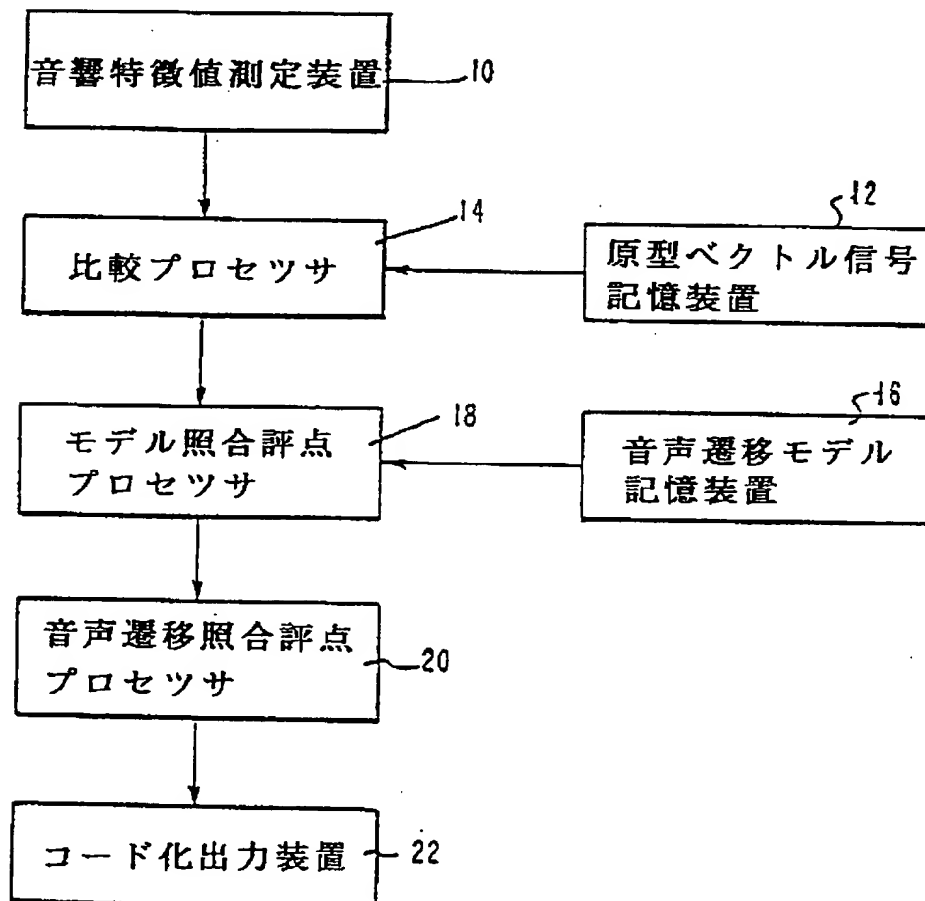


図1 本発明の音声コード化装置

【図2】

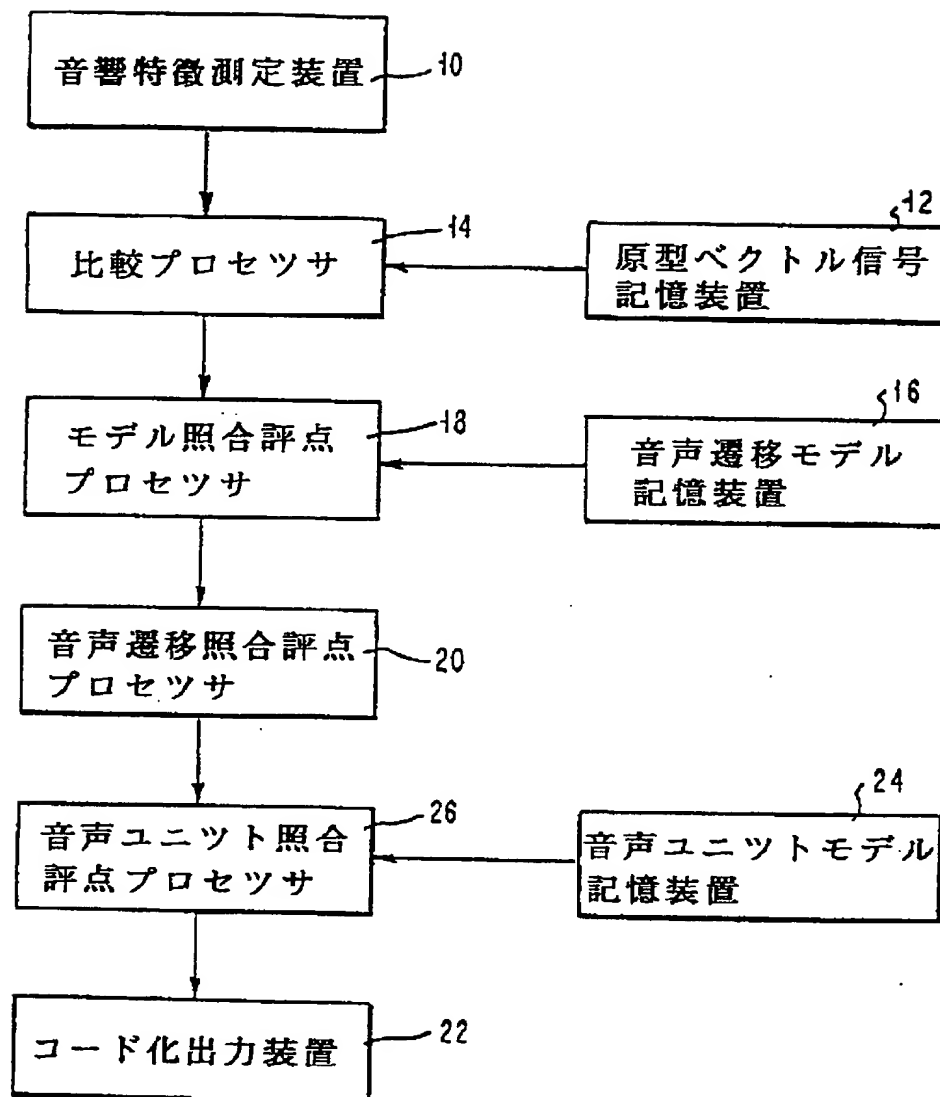


図2 本発明の他のコード化装置

【図3】

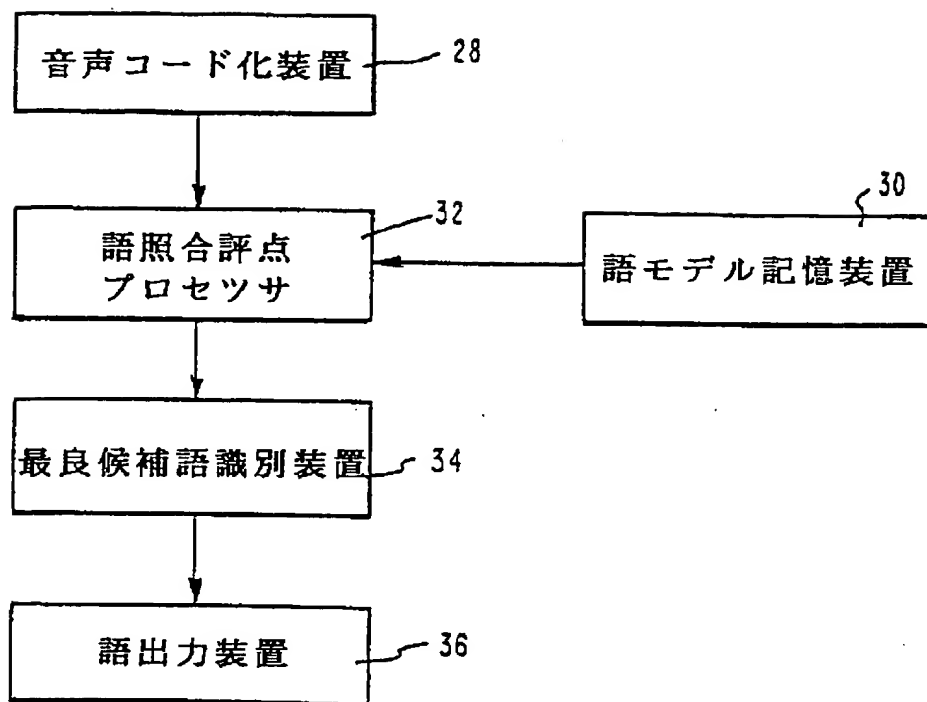


図3 本発明の音声コード化装置を用いた  
本発明の音声認識装置の一実施例

【図4】

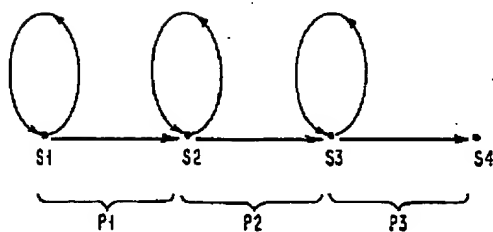


図4 語についての音響モデルの仮想の例

【図5】

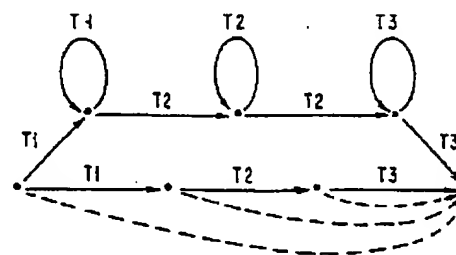


図5 音素についての音響モデルの仮想の例

【図6】

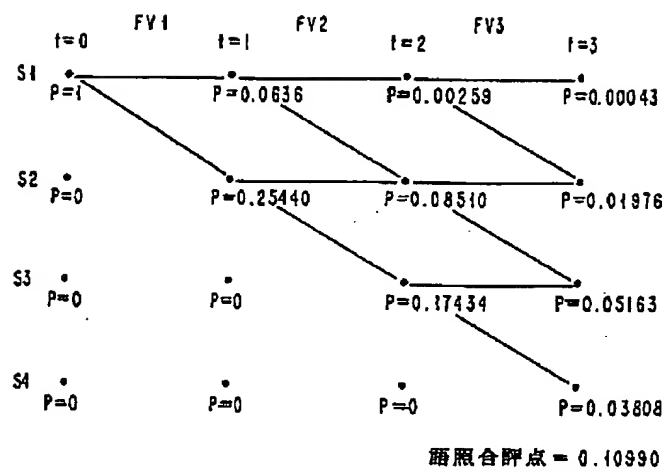


図6 図4の音響モデルを通る完全な経路  
及び一部の経路の仮想の例

【図6】

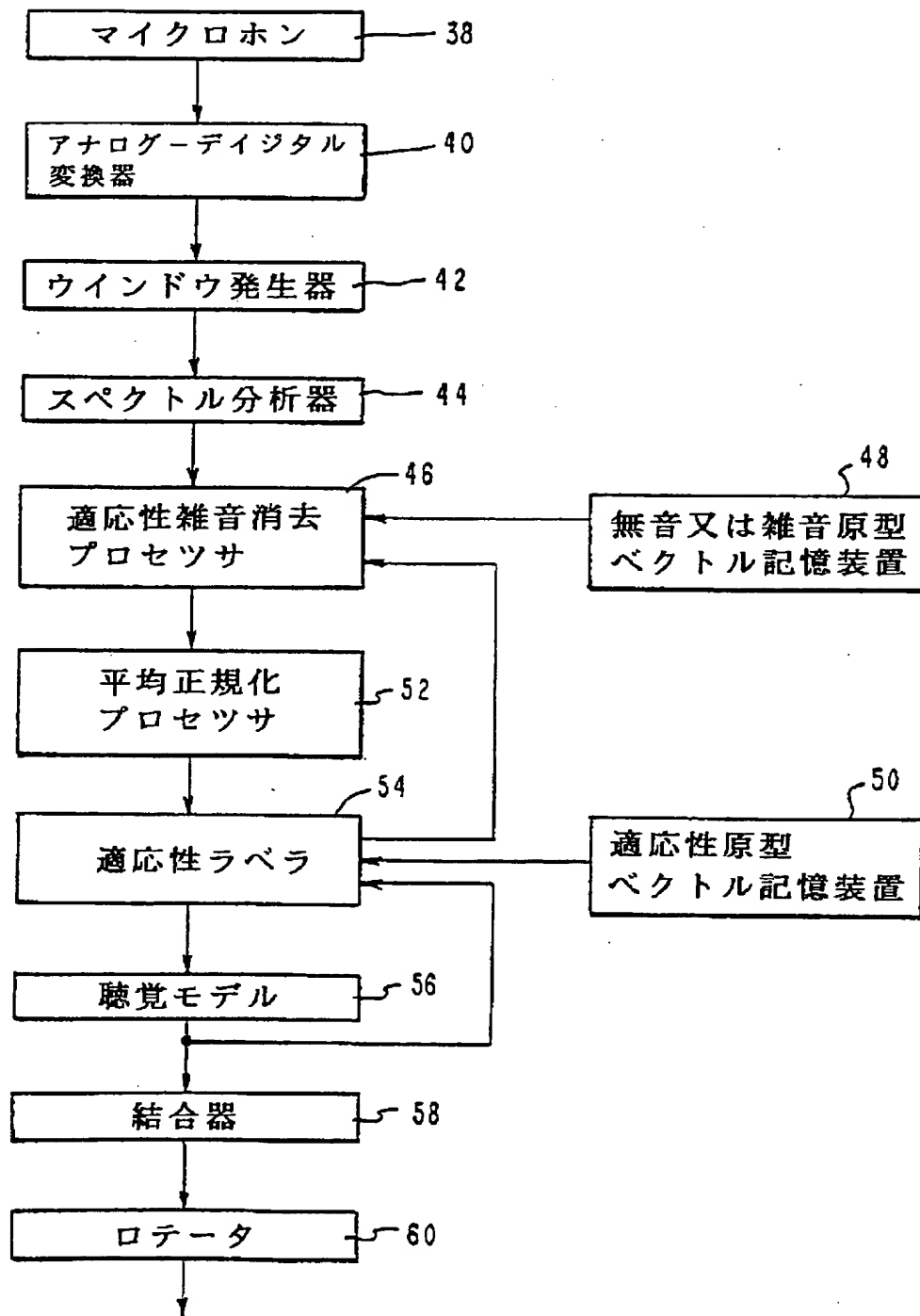


図7 音響特徴値測定装置の一実施例

## フロントページの続き

(72)発明者 ラリット・アール・パール  
アメリカ合衆国、ニューヨーク州10501、  
アマウオーク、エリシヤ・パーディ・ロー  
ド、ボックス28、アールデュー1 (番地な  
し)  
(72)発明者 ピーター・ビンセント・ドウソウザ  
アメリカ合衆国、カリフォルニア州95124、  
サン・ジヨウズ、ピスタ・ループ 6001番  
地

(72)発明者 ポナニ・エス・ゴバラクリシュナン  
アメリカ合衆国、ニューヨーク州10520、  
クロトナーオンーハドソン、シーニック・  
ドライブ 25ジエイ番地  
(72)発明者 マイケル・アラン・ピチエニー  
アメリカ合衆国、ニューヨーク州10404、  
ホワイト・プレーンズ、ラルフ・アベニュー  
118番地